IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	Examiner: Unassigned
	)	
Jacques COLINGE et al.	)	Group Art Unit: 2123
	)	
Serial No.: 10/624,531	)	Confirmation No.: 6658
	)	
Filed: July 23, 2003	)	Attorney Docket No.: 64176.000005

For: SYSTEM AND METHOD FOR SCORING PEPTIDE MATCHES

Mail Stop Petition
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

PETITION TO MAKE SPECIAL

This is a Petition to Make Special under 37 C.F.R. § 1.102(d) for accelerated examination of the above-referenced application that was filed on July 23, 2003. The petition fee under 37 C.F.R. § 1.17(h) is enclosed. The Commissioner is authorized to charge any further fees or credit any overpayments to Deposit Account No. 50-0206.

This Petition to Make Special is being made under MPEP 708.02 (VIII). In accordance with that section, Applicants have caused a pre-examination search to be made. A first part of the pre-examination search was made on the USPTO system EAST for U.S. and foreign patents and patent applications with emphasis on Class 250, subclasses 281, 282, 283, 286, 287, 288, 290, 292, and 339.07 using the following keywords, alone or in combination: "tandem," "spectrometry," "ratio," "variable," "random," "probability," "match," "peptide," "protein," and "stochastic." A second part of the pre-examination search was a database search on Google, Delphion, Patent Abstracts of Japan, Pubmed, and Esp@cenet. Examiner James Leybourne in

Class 250 (Art Unit 2881) was consulted (by Applicants' search agent) in confirming the field of search.

Applicants previously submitted a Supplemental Information Disclosure Statement (IDS) on January 24, 2005 submitting references cited in an International Search Report dated October 15, 2004. Another Supplemental IDS is being filed herewith disclosing additional references discovered in the pre-examination search.

The present utility patent application was filed on July 23, 2003 claiming priority to provisional application No. 60/468,580, filed on May 7, 2003, and provisional application No. 60/399,464, filed on July 29, 2002.

The pending claims in the present application recite a system and method for scoring peptide matches. The scoring of peptide matches is preferably based on tandem mass spectrometry (MS/MS) data.

According to independent claim 1, to score a match between a first peptide and a second peptide, a stochastic model may be generated based on one or more match characteristics associated with the first peptide, the second peptide and their fragments. A first probability that the first peptide matches the second peptide, and a second probability that the first peptide does not match the second peptide, may be calculated, each based on the stochastic model. A match between the first peptide and the second peptide may be scored based at least in part on a ratio between the first probability and the second probability. The ratio is referred to as a likelihood ratio.

More specifically, according to independent claim 12, an extended match E may be defined based on mass spectrum information associated with an experimental peptide and a candidate peptide. A stochastic model may be generated based on the mass spectrum

information. And the extended match E may be scored based on a likelihood ratio $L = \frac{\mathbf{P}(E|D,s,H_1)}{\mathbf{P}(E|D,s,H_0)}$, where D denotes extra information that is associated with the experimental peptide and the candidate peptide; s is a peptide sequence; H_1 is a hypothesis that the peptide sequence s is the correct sequence of the experimental peptide; H_0 is a null-hypothesis that the peptide sequence s is an erroneous sequence of the experimental peptide; and probabilities $\mathbf{P}(E|D,s,H_1)$ and $\mathbf{P}(E|D,s,H_0)$ are calculated based on the stochastic model.

Out of all of the references submitted to the Patent Office, the following are believed by the Applicants to be the most relevant to the claims pending in the application.

1. **U.S. Patent No. 5,538,897 to Yates, III et al ("Yates-1")**

Yates-1 discloses a method for correlating a peptide fragment mass spectrum with amino acid sequences derived from a database. A protein sequence database is used to predict candidate fragment spectra. The predicted fragment spectra are then compared with an experimentally-derived fragment spectrum to determine the best match or matches. Preferably, the parent peptide, from which the fragment spectrum was derived, has a known mass. Sub-sequences of the various sequences in the protein sequence database are analyzed to identify those sub-sequences corresponding to a peptide with same or similar mass as the parent peptide in the fragment spectrum. For each sub-sequence having the proper mass, a predicted fragment spectrum can be calculated, e.g., by calculating masses of various amino acid subsets of the candidate peptide. The result will be a plurality of candidate peptides, each with a predicted fragment spectrum. The predicted fragment spectra can then be compared with the experimentally-derived fragment spectrum using a closeness-of-fit measure, preferably

calculated with a two-step process, including a calculation of a preliminary score and, for the highest-scoring predicted spectra, calculation of a correlation function.

Yates-1 does not teach or suggest Applicants' invention because the Yates-1 scoring scheme is not a probability-based approach at all. Yates-1 does not score a match between the candidate fragment spectra and the experimentally-derived fragment spectrum based on any random variables or stochastic model. Nor does Yates-1 calculate any probabilities or likelihood ratio. Instead, the Yates-1 scoring scheme relies on a "closeness-of-fit score" and a "correlation function," neither of which involves probabilistic measures. See Yates-1, col. 6, Equations (1) and (2). Therefore, Applicants' invention is fundamentally different from Yates-1.

2. **U.S. Patent No. 6,017,693 to Yates, III et al ("Yates-2")**

(Yates-2 was cited in the Supplemental IDS filed on January 24, 2005.)

Yates-2 discloses a method for correlating a peptide fragment mass spectrum with amino acid sequences derived from a database. The scoring scheme disclosed in Yates-2 (subject to terminal disclosure) is substantially the same as the scheme disclosed in Yates-1. Yates-2 also relies on a closeness-of-fit measure and a correlation function to score matches for nucleotides, amino acids or carbohydrates. As discussed above, since its scoring scheme is essentially a heuristic algorithm rather than a probability-based method, Yates-2 does not teach or suggest Applicants' invention.

3. **U.S. Patent No. 6,393,367 to Tang et al. ("Tang")**

(Tang was cited in the Supplemental IDS filed on January 24, 2005.)

Tang discloses a method for determining the probability that a biological molecule identification is incorrect for a chosen significance level. The method comprises the steps of: a)

generating theoretical mass data for biological molecules; b) generating an experimental mass data for an unknown biological molecule; c) comparing the experimental mass data generated in step (b) with each theoretical mass data generated in step (a); d) calculating a score for each comparison in step (c), wherein the score is a function of the similarity between mass data compared; e) selecting at least two scores from the scores in step (d) to form a primary data set, wherein the scores correspond to a comparison that denotes a degree of similarity between the mass data compared; f) generating a sufficient quantity of artificial data sets from the primary data set in step (e); g) calculating a sample mean for each artificial data set in step (f); h) estimating population mean and population standard deviation from the sample means generated in step (g), wherein the population is based on the distribution underlying the primary dataset; i) computing a Z score from the population mean and population standard deviation for each score calculated in step (d) to standardize the scores; j) choosing a significance level; and k) comparing a test Z score to a Z score of the chosen significance level to determine the probability that the biological molecule identification is incorrect.

The comparison scores in Tang are generated based on then existing algorithms such as ProFound. Tang, col. 5, lines 60-62. And Tang only uses the comparison score as a measure of the degree of similarity between the theoretical and experimental mass data. The similarity is assessed by comparing every experimental mass with every theoretical mass. Tang, col. 6, lines 1-4 and lines 26-29. The comparison scores are not based on a likelihood ratio as recited in Applicants' invention. The Z score in Tang is calculated from artificial datasets created from the mass data comparisons, not from any stochastic model. Therefore, Tang does not teach or suggest the scoring method as recited in Applicants' invention.

4. **U.S. Patent No. 6,489,121 to Skilling ("Skilling-1")**

Skilling-1 discloses a method of identifying the most probable amino acid sequences which would account for the mass spectrum of a protein or peptide. The method models the fragmentation of a peptide or protein in a tandem mass spectrometer to facilitate comparison with an experimentally determined spectrum. A fragmentation model is used which takes account of all possible fragmentation pathways which a particular sequence of amino acids may undergo. A peptide or protein is identified by comparing an experimentally determined mass spectrum with spectra of trial sequences predicted using the fragmentation model from a library of known peptides or proteins. The fragmentation model sums probabilistically over all the ways in which a trial sequence might fragment and give rise to peaks in the experimentally determined mass spectrum.

Skilling-1 does not teach or suggest Applicants' invention. Skilling-1 scores a match with a trial sequence solely based on the probability of a fragmentation route that produces the trial sequence. Skilling-1 does not treat match characteristics as random variables in a stochastic model. Nor does Skilling-1 teach or suggest calculating a likelihood ratio that factors the chances for both a "hit" and a "miss" into scoring the peptide matches.

5. **U.S. Patent No. 6,489,608 to Skilling ("Skilling-2")**

Skilling-2 discloses a method for determining the sequence of amino acids that constitute peptides, polypeptides or proteins by mass spectrometry and especially by tandem mass spectrometry. The method comprises the steps of: producing a processable mass spectrum from a peptide; choosing a limited number of trial sequences of amino acids which are consistent with a prior probability distribution; and iteratively modifying the trial sequences through a

terminated Markov Chain Monte Carlo algorithm to generate new trial sequences of amino acids consistent with the prior probability distribution, using at each stage modifications which lie within the prior probability distribution, calculating the probability of each of the trial sequences accounting for the processable mass spectrum, and accepting or rejecting each of the trial sequences according to said calculated probability and the mathematical principle of detailed balance. The prior probability distribution was assigned to the trial sequences based on pseudo-random combinations of the amino acid residues comprised in a library and are probabilities that reflect the natural abundance of the amino acids concerned.

Skilling-2 does not teach or suggest Applicants' invention. Skilling-2 does not score an extended match using a stochastic model as recited in Applicants' invention. Skilling-2 only calculates the probability of a trial sequence accounting for the processable mass spectrum, but does not calculate the probability for the null-hypothesis that the trial sequence is erroneous. As a result, Skill-2 does not contemplate computing a likelihood ratio for peptide matches.

6. **U.S. Patent No. 6,582,965 to Townsend et al. ("Townsend")**

Townsend discloses a method for generating a library of peptides based on a peptide of a predetermined molecular mass and determining the amino acid sequence of the peptide from the library. The library is generated by defining a set of all allowed combinations of amino acids that can be present in the unknown peptide, where the molecular mass of each combination corresponds to the predetermined molecular mass within the experimental accuracy, and generating an allowed library of all possible permutations of the linear sequence of amino acids in each combination in the set.

Townsend adopts a 2-step scoring scheme that is essentially identical to what is disclosed in Yates-1 and Yates-2 above. That is, Townsend calculates a measure of closeness-of-fit between the predicted mass spectra and the experimentally-derived fragment spectra in two steps: calculating a preliminary closeness-of-fit score and then calculating a correlation function for the highest-scoring amino acid sequence. See Townsend, col. 8, lines 49-57 and col. 9-10. As discussed above, the measure of closeness-of-fit is not a probability-based method as recited in Applicants' invention.

7. **U.S. Patent No. 6,800,449 to Haynes et al. ("Haynes")**

Haynes discloses a method of identifying proteins with a shared function from a protein pool. The method comprises preparing a protein pool. The protein pool is applied to a functional affinity column wherein the functional affinity column isolates proteins with a common function based on the affinity chromatographic behavior of the proteins. The isolated proteins are analyzed using a one or more dimensional column in combination with mass spectrometry thereby producing spectral information. The isolated proteins are identified by matching the spectral information with a theoretical mass spectrum of a protein having a known sequence.

Haynes does not disclose any probability-based scoring method as recited in Applicants' invention. In Haynes, the focus is on isolating proteins based on shared functions and tandem MS scoring is just an auxiliary step for matching the isolated proteins to known sequences. Apart from the brief reference to SEQUEST and Xcorr scores (col. 14, lines 2-3 and lines 32-33), Haynes does not disclose any scoring methods in detail.

8. **U.S. Patent No. 6,852,544 to Aebersold et al. ("Aebersold")**

Aebersold discloses analytical reagents and mass spectrometry-based methods using these reagents for the rapid, and quantitative analysis of proteins or protein function in mixtures of proteins. Similar to the Haynes patent discussed above, Aebersold is focused on isolation of peptide fragments and does not disclose any scoring methods in detail.

9. **U.S. Patent Application (Pub. No. 2004/0041089) by Zhu et al. ("Zhu")**

Zhu discloses a method for locating pattern matches in amino acids by use of various and sequential filters capable of determining inner sample pattern matches, inner group pattern matches, and word matching for purposes of further analysis or data mining. Filters include the use of a scoring scheme, comparison of scan numbers versus sequence of common ions to be MS/MS, and daughter ion subtraction for obtaining pattern match candidates.

Zhu does not teach or suggest Applicants' invention. Zhu merely uses software bundled with MS instrument such as SEQUEST, Qstar and Sonar for word matching. The scoring system provides a cumulative score, a Q_ratio, a T_ratio, and a t_score, none of which resembles the likelihood ratio disclosed in Applicants' invention. Further, the Zhu application was filed on August 30, 2002, which was after the earliest priority date (July 29, 2002) of the present application.

10. **U.S. Patent Application (Pub. No. 2004/0044481) by Halpern ("Halpern")**

Halpern discloses a method for comparing a query peptide to a plurality of database peptides using mass spectrometry data from the query peptide and a pre-calculated peptide index. The method comprises the steps of: (a) constructing an index table comprising a plurality of peptide mass values using masses obtained from database peptides and backbone ion fragments

thereof; (b) identifying query mass values associated with the query peptide and query peptide backbone fragments or ions; (c) identifying query mass values that correspond to masses contained in the index table and generating comparison scores which reflect the correspondence between the query mass values and the masses contained in the index table; and (d) evaluating the comparison scores to identify database peptide related to the query peptide based upon the greatest comparison score.

In Halpern's scoring operation, a plurality of mass scores maintained for one or more peptides in the database are incremented based on the identification of retrieved entries within the index table having substantially the same associated mass. That is, the comparison score is essentially a count of database hits. See, e.g., paragraphs 64 and 66. Thus, Halpern's scoring operation does not follow a probabilistic approach as recited in Applicants' invention and Halpern does not contemplate the use of a likelihood ratio for peptide matches.

11. **U.S. Patent Application (Pub. No. 2004/0175838) by Jarman et al. ("Jarman")**

Jarman discloses a scoring method for peptide identification based on a probabilistic model for the occurrence of spectral peaks corresponding to key partial peptide ion types. In particular, the ion frequencies for the most frequently observed ion types are initially estimated from a training data set of known sequences. These frequencies are then used to construct a fingerprint for any candidate peptide of interest, where the fingerprint consists of a list of spectral peaks and their corresponding probabilities of appearance. A spectrum is then scored against the candidate fingerprints using a likelihood ratio between the hypothesis that the candidate peptide is not present and the hypothesis that the candidate peptide is present. This likelihood ratio can be used for peptide identification. In addition, a probabilistic score that estimates the probability

of a candidate peptide being present in the test sample can be constructed from the likelihood ratio.

It should be noted that the Jarman application was filed on February 10, 2003, which was after the earliest priority date (July 29, 2002) of the present application. Although Jarman discloses a method that uses a likelihood ratio for peptide identification, Jarman does not disclose an extended match or a stochastic model as recited in Applicants' invention.

12. **"SCOPE: A Probabilistic Model for Scoring Tandem Mass Spectra Against a Peptide Database" by Bafna et al. ("Bafna")**

(Bafna was cited in the Supplemental IDS filed on January 24, 2005.)

Bafna proposes a two-stage stochastic model for the process of MS/MS spectrum generation from a given a peptide. The first step involves generation of fragments from a peptide, according to a probability distribution estimated from many training samples. The second step involves the generation of a spectrum from the fragments according to the distribution of the instrument measurement error. The model explicitly incorporates fragment ion probabilities, noisy spectra, and instrument measurement error.

In Bafna, a peptide is scored only by the probability that the observed spectrum is generated by this peptide. Bafna does not consider the probability for the null-hypothesis. As a result, Bafna does not contemplate the use of a likelihood ratio in scoring the peptide matches. Further, Bafna's stochastic model does not exclude noise data as does Applicants' invention.

13. **"ProFound: An Expert System for Protein Identification Using Mass Spectrometric Peptide Mapping Information" by Zhang et al. ("Zhang")**

(Zhang was cited in the Supplemental IDS filed on January 24, 2005.)

Zhang describes the protein search engine "ProFound" which employs a Bayesian algorithm to identify proteins from protein databases using mass spectrometric peptide mapping data. The ProFound algorithm is essentially a peptide mass fingerprinting (PMF) technique which relies on the set of masses of peptide fragments produced by cleavage of the protein by an enzyme of high cleavage specificity. Applicants' invention is related to tandem mass spectrometry, which is different from the PMF technique. Furthermore, Zhang only calculates a probability for the hypothesis that "protein *k* is the protein being analyzed" based on the assumption that the protein being analyzed exists in the database. Thus, Zhang does not teach or suggest calculating the probability for the null-hypothesis or using a likelihood ratio to score a match.

14. **"Database Searching Using Mass Spectrometry Data" by Yates, III ("Yates")**

(Yates was cited in the Supplemental IDS filed on January 24, 2005.)

Yates provides a general survey of using mass spectrometry in conjunction with database searching. The paper does not disclose new tandem mass spectrometry methods but only discusses existing techniques at that time. The SEQUEST, FASTA, and EST databases searches mentioned in the paper do not teach or suggest a scoring method wherein a likelihood ratio is calculated for peptide matches. Therefore, Yates can neither anticipate Applicants' invention nor render it obvious.

15. **"An Alternative to the SEQUEST Cross-Correlatio Scoring Algorithm for Tandem Mass Spectral Identification Through Database Lookup: the Luck Scoring Function, and the Probability of an Unrelated Spectra Match Model" by Fridman et al. ("Fridman)**

Fridman discloses a new method with high discriminating power for searching protein sequence databases for peptide identification. Fridman develops a model to derive the

probability distribution of degree of match between an experimental spectrum a theoretical spectrum from a database peptide, assuming that the theoretical spectrum is produced by a different unrelated peptide. Based on this probability distribution, a *Luck* score is calculated for the match between each experimental-theoretical spectral pair.

Though the Fridman method is a probability-based approach, it does teach or suggest calculating a likelihood ratio as a score for peptide matches. Also, based on the 2003 date of its first reference, Fridman was published after the earliest priority data (July 29, 2002) of the present application.

In summary, the prior art discovered by the Applicants during the pre-examination search all fail to disclose a stochastic model for an extended peptide match that takes into account characteristics associated with the peptides and their fragments. Nor does any prior art reference teach or suggest calculating a likelihood ratio to score peptide matches. The prior art references fail to show or suggest a method or system as described and claimed by the present invention.

On the basis of the foregoing, the Applicant respectfully requests granting this Petition To Make Special so that the application will be taken up promptly.

Respectfully submitted,

HUNTON & WILLIAMS LLP

Dated: 4/5/05

By: _____

Ce Li
Limited Recognition
Under 37 C.F.R. § 10.9(b)

Hunton & Williams LLP
Intellectual Property Department
1900 K Street, N.W.
Suite 1200
Washington, DC 20006
(202) 955-1500 (telephone)
(202) 778-2201 (facsimile)